



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### **Molecular genetic aetiology of general cognitive function is enriched in evolutionarily conserved regions**

**Citation for published version:**

neuroCHARGE Cognitive Working group, Hill, WD, Davies, G, Harris, S, Hagenaars, SP, Liewald, DC, Penke, L, Gale, CR & Deary, IJ 2016, 'Molecular genetic aetiology of general cognitive function is enriched in evolutionarily conserved regions', *Translational Psychiatry*, vol. 6, no. 12, e980.  
<https://doi.org/10.1038/tp.2016.246>

**Digital Object Identifier (DOI):**

[10.1038/tp.2016.246](https://doi.org/10.1038/tp.2016.246)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Translational Psychiatry

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



## ORIGINAL ARTICLE

## Molecular genetic aetiology of general cognitive function is enriched in evolutionarily conserved regions

WD Hill<sup>1,2</sup>, G Davies<sup>1,2</sup>, SE Harris<sup>1,3</sup>, SP Hagenaars<sup>1,2,4</sup>, The neuroCHARGE Cognitive Working group<sup>8</sup>, DC Liewald<sup>1,2</sup>, L Penke<sup>1,5,6</sup>, CR Gale<sup>1,2,7</sup> and IJ Deary<sup>1,2</sup>

Differences in general cognitive function have been shown to be partly heritable and to show genetic correlations with several psychiatric and physical disease states. However, to date, few single-nucleotide polymorphisms (SNPs) have demonstrated genome-wide significance, hampering efforts aimed at determining which genetic variants are most important for cognitive function and which regions drive the genetic associations between cognitive function and disease states. Here, we combine multiple large genome-wide association study (GWAS) data sets, from the CHARGE cognitive consortium ( $n = 53\,949$ ) and UK Biobank ( $n = 36\,035$ ), to partition the genome into 52 functional annotations and an additional 10 annotations describing tissue-specific histone marks. Using stratified linkage disequilibrium score regression we show that, in two measures of cognitive function, SNPs associated with cognitive function cluster in regions of the genome that are under evolutionary negative selective pressure. These conserved regions contained  $\sim 2.6\%$  of the SNPs from each GWAS but accounted for  $\sim 40\%$  of the SNP-based heritability. The results suggest that the search for causal variants associated with cognitive function, and those variants that exert a pleiotropic effect between cognitive function and health, will be facilitated by examining these enriched regions.

*Translational Psychiatry* (2016) 6, e980; doi:10.1038/tp.2016.246; published online 13 December 2016

## INTRODUCTION

Individual tests of cognitive function correlate positively, allowing a single latent factor to be extracted from a battery of tests.<sup>1</sup> This general cognitive factor typically accounts for around 40% of the phenotypic variation in a battery of mental tests and, in large molecular genetic studies, has been shown to be heritable with common genetic variants in total explaining around 30% of phenotypic variation.<sup>2–4</sup> A higher level of general cognitive function is associated with better health across a range of diseases, both psychiatric and physical, and with lower all-cause mortality.<sup>5</sup> More recently these phenotypic associations between general cognitive function<sup>6</sup> and individual tests of cognitive function<sup>7</sup> with health have been shown to be partly the result of genetic correlations, indicating pleiotropy, meaning that health states show positive correlations with cognitive function in part because the same genetic variants are associated with both cognitive function and health.<sup>6</sup> However, although general cognitive function, along with performance on individual tests of cognitive function, is known to be heritable and to exhibit genetic correlations with health states, for cognitive function, few loci have attained genome-wide statistical significance.<sup>4,8</sup> This hampers the effort to understand how genetic variation can result in individual differences in general cognitive function and in turn how this is also associated with variation in health.

This large difference between the variance explained by single-nucleotide polymorphisms (SNPs) that do reach genome-wide

significance and heritability estimates derived using all tested SNPs, indicates that much of the heritability of general cognitive function lies in SNPs that have not attained genome-wide significance. Although an increase in sample size will result in an increase in the power to detect significant effects of SNPs in a genome-wide association study (GWAS),<sup>9</sup> the problem remains of organizing these individual hits into a coherent description of the genetic architecture of cognitive function. Another method that can both increase statistical power and facilitates an understanding of how genetic variation can result in phenotypic variation is gene set analysis (GSA).<sup>10</sup> GSA tests the hypothesis that a set of genes, united by a shared biological function<sup>11</sup> or their previous association with another phenotype,<sup>12</sup> jointly show an association with the phenotype of interest. The GSA method exploits phenotypes where a highly polygenic architecture is evident by summing the small effects of multiple variants located within the predefined gene set. GSA is not reliant on any single variant attaining genome-wide significance. As such, statistical power is increased as the number of statistical tests is reduced and individual weak effects are combined together to produce a stronger association signal.<sup>12,13</sup>

Multiple methods exist for the analysis of groups of SNPs treated as the unit of association.<sup>10,14</sup> However, many methods assume only a single causal SNP in each of the genetic loci,<sup>15</sup> which, along with being more likely to inflate the type 1 error rate,<sup>16</sup> also fails to model a polygenic architecture. Other methods suffer from limitations such as requiring access to participants'

<sup>1</sup>Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK; <sup>2</sup>Department of Psychology, University of Edinburgh, Edinburgh, UK; <sup>3</sup>Medical Genetics Section, University of Edinburgh Centre for Genomics and Experimental Medicine and MRC Institute of Genetics and Molecular Medicine, Western General Hospital, Edinburgh, UK; <sup>4</sup>Division of Psychiatry, University of Edinburgh, Edinburgh, UK; <sup>5</sup>Georg Elias Müller Institute of Psychology, Georg August University Göttingen, Göttingen, Germany; <sup>6</sup>Leibniz Science Campus Primate Cognition, Göttingen, Germany and <sup>7</sup>MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, UK. Correspondence: Professor IJ Deary, Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK.

E-mail: ian.deary@ed.ac.uk

<sup>8</sup>The members of the neuroCHARGE Cognitive Working group are listed before References.

Received 30 July 2016; revised 4 October 2016; accepted 17 October 2016

genotypes,<sup>17</sup> and other methods fail to account for linkage disequilibrium (LD) that can lead to the same SNP being counted multiple times within a single gene set.<sup>18</sup> Here, we make use of a recently developed method, stratified linkage disequilibrium score regression,<sup>19</sup> that requires access to only genome-wide association summary level data. This method utilizes information from each SNP in a functional category while explicitly modelling LD to show whether a category is associated with a greater proportion of the heritability of a trait than the proportion of SNPs it contains would suggest. We apply this method to the current largest GWAS on general cognitive function.<sup>4</sup> We also examine specific tests of cognitive function using the UK Biobank data set. We search for an enrichment in the heritability found in 52 regions corresponding to functional annotations from across cell types, and 10 corresponding to cell-specific functional groups (see Supplementary Table 1). We examine these functional categories, because the distribution of significantly associated SNPs from hundreds of GWAS have indicated that, across diverse phenotypes, significant associations are more likely to be found in regulatory regions such as DNase hypersensitivity sites,<sup>20</sup> as well as in protein coding regions<sup>21</sup> and untranslated regions.<sup>22</sup> DNase hypersensitivity sites are regions of chromatin vulnerable to the DNase 1 enzyme, as the chromatin in these regions has lost its condensed structure and leaves the DNA exposed, whereas untranslated regions are involved in the regulation of translation of RNA. In addition, the role of evolutionarily conserved regions has been shown for disease states and psychiatric disorders,<sup>19</sup> many of which show genetic correlations with the individual tests of cognitive function used here.<sup>7</sup> By examining the contributions of each of these functional genetic categories, we aim to find regions of the genome that have relatively prominent roles in individual differences in general cognitive function.

## MATERIALS AND METHODS

### Samples

The data used for this study were the summary of GWAS statistics from the CHARGE consortium study on general cognitive function in middle and older age, which had a total of 53 949 individuals,<sup>4</sup> and a study of verbal-numerical reasoning (VNR) based on the UK Biobank<sup>8</sup> with 36 035 individuals. We next derived a heritability Z-score for both of the data sets that was defined as the heritability estimate produced by LDS regression divided by its standard error. The magnitude of the heritability Z-score is affected by three properties, sample size, SNP-based heritability and the proportion of causal variants. An increase in these three properties is associated with an increase in the heritability Z-score. This indicates that the heritability Z-score is capturing information about the genetic architecture of a trait, with traits that have sufficient power, from sample size, high heritability and a high proportion of causal variants yield the greatest heritability Z-scores.

Here, a heritability Z-score  $> 7$ , as used in Finucane *et al.*<sup>19</sup> was used as evidence for a sufficient polygenic signal within the data set for use with stratified LD regression. The CHARGE data set yielded a heritability Z-score of 10.54 and in the VNR test a heritability Z-score of 9.64 was derived. This indicates that both of these data sets have sufficient power for use with stratified LDS regression.

### Cognitive phenotypes

**General cognitive function.** The CHARGE cognitive working group published a GWAS of general cognitive function in 53 949 middle and older age adults.<sup>4</sup> General cognitive function describes the statistically revealed overlap between tests of cognitive function, that is, people who do well on one type of cognitive test tend to do well on others. The cognitive tests included in the general cognitive components used by the CHARGE cognitive working group's contributing cohorts generally measure fluid cognitive functions. These are functions assessed by tests that tend to include unfamiliar materials, that do not draw upon a participant's level of general knowledge, and that tend to show a negative trend with age. Each of the CHARGE GWAS project's cohorts used a different battery of mental tests. Full details of the tests used to measure general fluid cognitive

function in each of the CHARGE consortium's cohorts can be found in Davies *et al.*<sup>4</sup>

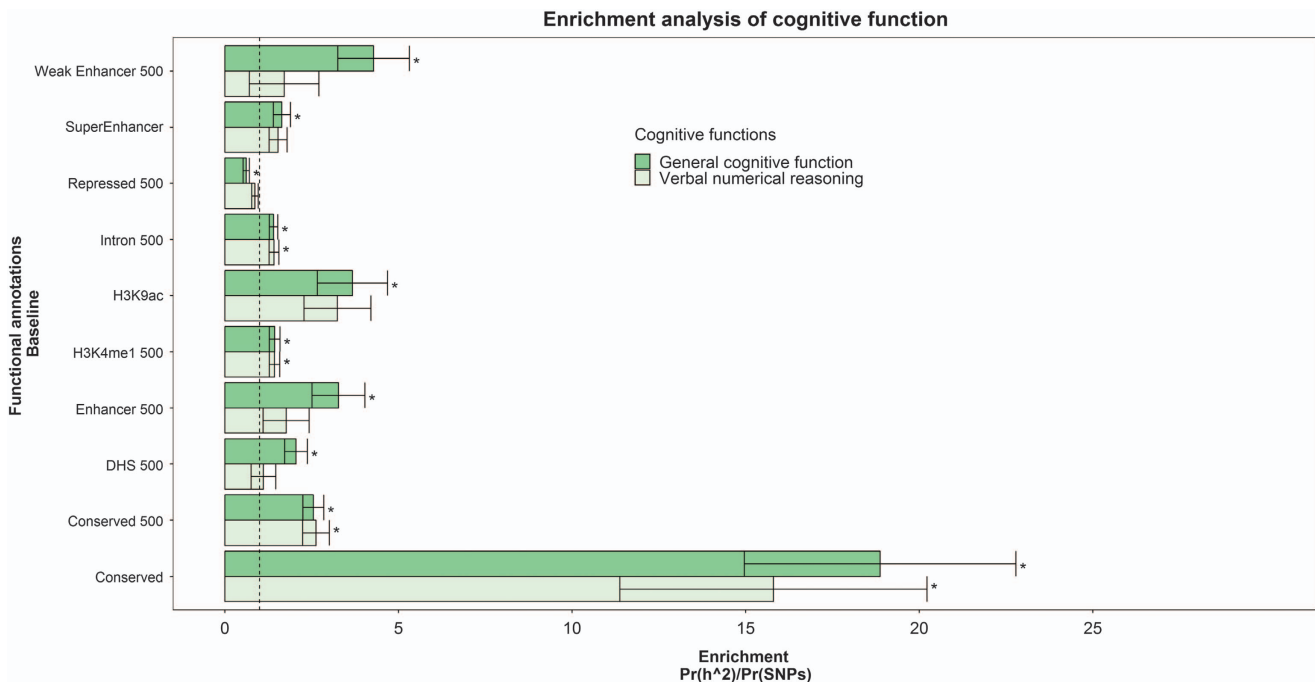
**Verbal-numerical reasoning.** The VNR test in UK Biobank consists of a series of 13 multiple choice questions that are answered in a 2 min time period. Six of the items were verbal questions and the remaining seven were numerical. An example of a verbal question is 'Bud is to flower as child is to?' (Possible answers: 'Grow/Develop/Improve/Adult/Old/Do not know/Prefer not to answer'). An example numerical question is 'If sixty is more than half of seventy-five, multiply twenty-three by three. If not subtract 15 from eighty-five. Is the answer?' (Possible answers: '68/69/70/71/72/Do not know/Prefer not to answer'). To some extent, these questions draw upon materials and information that the participants should be familiar with and the scores on this test are stable when comparing the means between the ages of 40 and 60 years with a linear decline evident from a comparison between the ages of 60 and 70. Genotype data were available from 36 035 individuals who had completed this test. Full details of the genotyping procedures used for this phenotype can be found in Davies *et al.*<sup>8</sup>

Although a composite measure of general cognitive ability would have been preferable to the use of a single test, some of the cognitive tests used in UK Biobank are of poor quality. The tests of memory and reaction time used in our previous Genome-Wide-Association Study (GWAS)<sup>8</sup> had test-retest correlations of 0.15 ( $n = 19\,872$ ) and 0.54 ( $n = 20\,188$ ), respectively; in addition, the memory score was based on one  $3 \times 2$  grid, and the reaction time test on just four trials. In addition, neither of these two tests showed significant genetic correlations with a general factor of cognitive ability constructed using validated tests (memory,  $r_g = 0.100$ ,  $s.e. = 0.112$ ,  $P = 0.370$ , reaction time,  $r_g = 0.067$ ,  $s.e. = 0.089$ ,  $P = 0.451$ ).<sup>7</sup> The VNR test, which we used here, had 13 items, shows a greater level of test-retest reliability ( $r = 0.65$ ) and had a very high genetic correlation with a general factor of cognitive ability ( $r_g = 0.812$ ,  $s.e. = 0.094$ ,  $P = 6.2 \times 10^{-18}$ ). On the basis of these findings we decided, *a priori*, not to include the memory or reaction time tests from our previous GWAS as they are unreliable and appear to have a very different genetic architecture from established indicators of general cognitive function.

### Statistical analysis

**Genetic correlations between phenotypes.** Owing to the phenotypic overlap between tests of cognitive function,<sup>1</sup> we first examine the degree to which the VNR measure from UK Biobank overlaps genetically with general cognitive function. Genetic correlations were derived using the summary statistics from general cognitive function in CHARGE and VNR in UK Biobank sets using LD score regression.<sup>23</sup> The same data processing pipeline was used here as by Bulik-Sullivan *et al.*<sup>23</sup> where a minor allele frequency of  $> 0.01$  was used and only those SNPs found in the HapMap3 with 1000 Genomes EUR with a minor allele frequency of  $> 0.5$  were retained. The integrated\_phase1\_v3.20101123 was used for LDS regression. Also, Indels, structural variants and strand-ambiguous SNPs were removed. Genome-wide significant SNPs were removed, as well as SNPs with effect sizes of  $\chi^2 > 80$ , as the presence of outliers can increase the standard error in a regression model. LD scores and weights for use with the GWAS of European ancestry were downloaded from the Broad Institute ([http://www.broadinstitute.org/~bulik/eur\\_ldscores/](http://www.broadinstitute.org/~bulik/eur_ldscores/)).

**Partitioned heritability.** General cognitive function<sup>4</sup> and VNR<sup>8</sup> were analysed using stratified LD score regression, where we followed the data processing steps of Finucane *et al.*<sup>19</sup> Stratified LD score regression belongs to a class of techniques that exploit the correlated nature of SNPs. Using this method, heritability can be derived by regressing SNP's association statistic (converted to a Z-score) onto its LD score. An LD score is the sum of squared correlations between the minor allele count of a SNP, with the minor allele count of every other SNP. As the correlational structure of the genome (LD) is used in deriving the heritability estimate, LD is controlled for. A full description of how this method works can be found in Finucane *et al.*<sup>19</sup> This heritability estimate is then used to derive an enrichment metric defined as the proportion of heritability captured by the functional annotation over the proportion of SNPs contained within it,  $(Pr(h^2)/Pr(SNPs))$ . This ratio describes whether a functional annotation contains a greater or lesser proportion of the heritability than would be predicted by the proportion of SNPs it contains,  $Pr(h^2)/Pr(SNPs) = 1$ . The proportion of the heritability for each category is used as the numerator, rather than the heritability of each category. This is due to Genomic Control (GC) being performed on most GWAS data sets and, as a result, the attenuation of the



**Figure 1.** A comparison between the functional annotations that were significantly enriched for general cognitive function. Enrichment was also found in evolutionarily conserved regions for verbal-numerical reasoning. Significant enrichment was also found across the phenotypes for single-nucleotide polymorphisms (SNPs) within 500 bp of introns and within 500 bp of the H3K4me1 histone mark. The enrichment statistic is the proportion of heritability found in each functional group divided by the proportion of SNPs in each group ( $\text{Pr}(h^2)/\text{Pr}(\text{SNPs})$ ). The dashed line indicates no enrichment found when  $\text{Pr}(h^2)/\text{Pr}(\text{SNPs}) = 1$ . Statistical significance is indicated with asterisk.

heritability estimate affects the total heritability and the heritability of each SNP set equally. As these are biased in the same direction and by the same amount, the proportion of heritability accounted for by each SNP set remains unaffected by the GC correction although the absolute heritability may change. Stratified LD scores were calculated from the European ancestry samples in the 1000 Genomes project (1000G) and only included the HapMap3 SNPs with a minor allele frequency of  $> 0.05$ .

The same functional annotations as those reported in Finucane *et al.*<sup>19</sup> were used. First, SNPs were assigned to a set of 24 overlapping publicly available functional annotations. Supplementary Table 1 details the full set of these functional categories, as well as the references used to construct them. An additional 500 bp window was placed around these annotations to prevent estimates being biased upwards by capturing enrichment in regions located close to the functional annotations.<sup>24</sup> A 100 bp window was also placed around chromatin immunoprecipitation and sequencing (ChIP-seq) peaks; the inclusion of these additional four sets resulted in a total of 52 overlapping functional SNP annotations which formed our baseline model. In addition, a further 10 sets were examined. These sets consisted of 220 cell-type-specific annotations for four histone marks (H3K4me1, H3K4me3, H3K9ac and H3K27ac), which were arranged into 10 broad categories corresponding to histone marks found in the central nervous system, immune and hematopoietic, adrenal/pancreas, cardiovascular, connective tissue, gastrointestinal, kidney, liver, skeletal muscle and other. The SNP sets examined here are not independent and the same SNP can appear in many of the sets examined here. The size of each of the SNP sets can be found in Supplementary Tables 2 and 3.

The 10 broad cell-type categories were then analysed by adding each of them to the full baseline model. This resulted in 10 additional tests, which included the baseline model and one of the 10 cell-type-specific groupings. In this way, enrichment for these cell-specific annotations was not driven by their also being a part of the baseline model. Multiple testing was controlled for by using false discovery rate correction applied to these 10 tests.

## RESULTS

### Genetic correlation

LDS regression was first used to determine the degree of overlap between the two cognitive phenotypes used. The genetic

correlation between general cognitive function and VNR was  $r_g = 0.783$ ,  $\text{s.e.} = 0.056$ ,  $P = 4.63 \times 10^{-45}$  indicating that many of the same genetic variants are involved in both these traits.

### Partitioned heritability

**General cognitive function in the CHARGE consortium.** Significant enrichment was found for 10 of the 52 functional annotations (Supplementary Figure 1 and Supplementary Table 2). Consistent with many quantitative traits,<sup>19,25</sup> SNPs that are found in evolutionarily conserved regions showed a high level of enrichment, where 2.5% of the SNPs accounted for 49.2% of the total heritability yielding an enrichment metric defined as  $\text{Pr}(h^2)/\text{Pr}(\text{SNPs})$  of 18.87 ( $\text{s.e.} = 3.91$ ),  $P = 4.88 \times 10^{-6}$ . Note that this does not mean that the 3% of the genome accounts for 40% of the total phenotypic variance; rather, it means that, of the 30% of phenotypic variance accounted for by common SNPs, the genetic variation in the ~3% of the genome accounts for 40% of that 30%. Statistically significant enrichment was also found after a 500 bp boundary was set around these regions (Figure 1).

Enrichment was also found for two of the histone marks, H3K9ac, where 46.3% of the heritability was found for 12.6% of SNPs (enrichment metric = 3.68,  $\text{s.e.} = 1.01$ ,  $P = 0.008$ ), and within 500 bp of H3K4me1, where 60.9% of the SNPs collectively explained 87.5% of the total heritability (enrichment metric = 1.44,  $\text{s.e.} = 0.15$ ,  $P = 0.004$ ). SNPs located within 500 bp of repressed regions showed a significant reduction in the level of heritability they captured. These regions accounted for 71.9% of the SNPs, but only explained 44.3% of the heritability (enrichment metric = 0.62,  $\text{s.e.} = 0.09$ ,  $P = 2.10 \times 10^{-5}$ ).

Statistically significant enrichment was also found for SNPs within 500 bp of weak enhancer regions which comprised 8.9% of the SNPs, that collectively explained 38.1% of the heritability of general cognitive function (enrichment metric = 4.28,  $\text{s.e.} = 1.03$ ,  $P = 0.001$ ). SNPs within 500 bp of the functional category of DNase hypersensitivity sites also demonstrated significant enrichment for



general cognitive function (enrichment metric = 2.05, s.e. = 0.33). These regions accounted for 49.9% of the SNPs but captured 100% of the heritability (s.e. = 16%). Although this does appear to be capturing the sum of the heritability present, it is not clear whether this is biologically meaningful as the inclusion of SNPs within 500 bp of this category raises the proportion of SNPs in the category from 17% to 50% indicating that the majority of the SNPs within the larger set are not at DNase hypersensitivity site. SNPs found within 500 bp of introns were also significantly enriched, accounting for 39% of the SNPs and 56% of the heritability of general cognitive function,  $P = 7.45 \times 10^{-4}$ .

The results for the cell type enrichment analysis indicated that histones that are marked specifically in cell types of the central nervous system accounted for 14.9% of the SNPs, but 45.0% of the heritability (enrichment metric = 3.03, s.e. = 0.51,  $P = 6.37 \times 10^{-5}$ ). The results for the 10 tissue types can be seen in Supplementary Figure 2. The full results for general cognitive function can be found in Supplementary Table 2.

**VNR in the UK Biobank sample.** The pattern of enrichment followed the same trend for VNR as for general cognitive function. Four of the five functional annotations found to be significantly enriched in general cognitive function were also enriched for VNR (Supplementary Figure 3 and Supplementary Table 3). For the baseline model, evolutionarily conserved SNPs were found to explain an enriched proportion of the heritability; 2.6% of SNPs were found to explain 41.2% of the heritability (enrichment metric = 15.80, s.e. = 4.42,  $P = 8.19 \times 10^{-4}$ ). As was found for general cognitive function, SNPs within 500 bp of introns also showed enrichment for heritability (enrichment metric = 1.2, s.e. = 0.14,  $P = 0.003$ ). This category contained 39.7% of the SNP that explained 56.2% of the heritability. Unlike general cognitive function, SNPs within 500 bp of the histone mark H3K9ac showed significant enrichment, rather than only SNPs found within this annotation (enrichment metric 2.5, s.e. = 0.48,  $P = 0.002$ ). This category contained 23.1% of the SNPs, which explained 58.1% of the heritability for VNR in the UK Biobank data set.

As was found in general cognitive function, histone marks specifically expressed in the central nervous system were found to contain a greater proportion of heritability (enrichment metric = 3.53, s.e. = 0.60,  $P = 2.40 \times 10^{-5}$ ). The enrichment results for each of the 10 tissue types can be seen in Supplementary Figure 4 and the full results for VNR can be found in Supplementary Table 3.

Figure 1 illustrates the significant annotations for general cognitive function and provides a comparison for how well these regions were enriched for VNR.

## DISCUSSION

We partitioned the total heritability found in two large, genetically correlated ( $r_g = 0.783$ , s.e. = 0.056) GWAS data sets on cognitive function into 24 broad functional annotations and 10 tissue types. Our analysis modelled LD, and took into account overlapping categories, as well as the proportion of SNPs in each category. We make a number of contributions to understanding the genetic architecture of cognitive function.

We find, for both of the cognitive phenotypes examined, the most substantial and statistically significant effects occurred in regions of the genome that are evolutionarily conserved in mammals. The SNPs within these regions accumulate the base-pair substitutions that differentiate species at a lower rate than would be expected under models of neutral selective pressure and these regions are depleted for the number of SNPs compared with regions that are not conserved. This indicates that a large portion of the common variants that are associated with cognitive function are under negative selective pressure. That 40% of the genetic variance in general cognitive function is under negative selection does not imply that higher cognitive function is

evolutionarily selected for, but that genetic variance that disrupts the evolutionarily old adaptive design encoded in these regions, thereby decreasing healthy cognitive function, is selected against. This supports the idea that mutation-selection balance has a substantial role in the genetics of general cognitive function,<sup>26</sup> particularly when mutational variation is introduced in evolutionarily conserved regions.

Evolutionarily conserved regions are unlikely to be specific to cognitive function, but rather to underlie fundamental design features important for general phenotypic functioning. The evolutionarily conserved regions used in the current paper have also been examined for enrichment with several disease- and health-related phenotypes. Significant enrichment was found for body mass index, schizophrenia and high-density lipoprotein cholesterol, but not for coronary artery disease, type 2 diabetes, low-density lipoprotein cholesterol or bipolar disorder.<sup>19</sup> The diseases and traits that were enriched in these conserved regions each show a genetic correlation with general cognitive function and individual tests of cognitive function (see refs 6,7 and Hill *et al.* (in prep)), whereas those that showed no enrichment at evolutionarily conserved regions were not genetically correlated with general cognitive function or the VNR test used here.<sup>7</sup> This suggests that not only do these evolutionarily conserved regions of the genome have a greater role in cognitive functions, but they may also harbour variants with pleiotropic effects on cognitive function, health and anthropometric traits, thereby reducing what has been variously called system integrity, developmental stability or general evolutionary fitness.<sup>26,27</sup>

Two previous studies using GSA have found that gene sets that are conserved between species are enriched for cognitive functions. A study by Hill *et al.*<sup>11</sup> found that common SNPs in the *N*-methyl-D-aspartate receptor complex were enriched for general cognitive function in two independent groups. The *N*-methyl-D-aspartate receptor complex is a component within the postsynaptic density (PSD), and using comparative proteomic analysis of the human and mouse PSD it has been found that the molecular composition of the postsynaptic density was highly similar, with more than 70% of the proteins found in the human PSD being found in the mouse PSD<sup>28</sup> indicating conservation between species. A high level of conservation has also been found between the proteins of the PSD in comparisons between human and chimp (last common ancestor (LCA) ~6 million years ago), as well as between mouse and rat (LCA 20 million years ago) and between human and mouse (LCA 90 million years ago), indicating conservation or negative selection indicative of conservation across the mammalian line.<sup>29</sup>

More recently, Johnson *et al.*<sup>30</sup> used the weighted gene co-expression network analysis to identify a novel module named M3 using cortical brain tissue extracted from living humans during surgery. This module was also present in both disease-free humans and in wild-type mouse hippocampi, indicating it had been conserved between both species. In addition, this module was found to be enriched for SNPs associated with general cognitive function and memory in two independent samples. The M3 also mapped poorly onto known biology, including the PSD, differentiating it from the *N*-methyl-D-aspartate receptor complex finding. In the current paper, we extend the findings of Hill *et al.*<sup>11</sup> and Johnson *et al.*<sup>30</sup> by considering all the SNPs, not just those found within genes, and show that regions of the genome that are under negative selective pressure harbour an enriched proportion of the heritable variance for cognitive function.

The gene sets used by Hill *et al.*<sup>11</sup> and Johnson *et al.*<sup>30</sup> were constructed to test specific synaptic components and networks of genes that work together. However, both of these gene sets are under selective pressure, the current paper expands on the findings of Hill *et al.*<sup>11</sup> and Johnson *et al.*<sup>30</sup> by showing that conserved regions, not just those found within genes, are enriched for cognitive function.

These results are also consistent with both theoretical predictions,<sup>31,32</sup> and empirical findings<sup>33</sup> of directional dominance acting on cognitive function. Directional selection is expected in traits that are related to fitness, and indicates a systematic direction of effect across causal loci. By examining runs of homozygosity, directional dominance can be seen when the phenotype of individuals that are homozygous for multiple locations differs from those that are heterozygous. Joshi *et al.*<sup>33</sup> examined runs of homozygosity in 354 224 individuals who had taken tests of cognitive ability and found that as homozygosity increased, scores on tests of cognitive ability went down, indicating selective pressure acting to increase levels of cognitive function. The present study expands on the works of Joshi *et al.*<sup>33</sup> by providing evidence that the regions of the genome most sensitive to genetic variation are more likely to harbour the variants that make the greatest contributions to cognitive and health differences.

SNPs within 500 bp of introns also showed significant enrichment for general cognitive function and for VNR. Enrichment for this region may indicate that, although not being translated into proteins, these regions may still exert an influence on individual differences in cognitive function. Indeed, Marioni *et al.*<sup>34</sup> have suggested that intronic regions are more likely to harbour genetic variation associated with normal cognitive function than exomic regions. In addition, SNPs within 500 bp of the H3K4Me1 histone mark were enriched across both cognitive phenotypes. General cognitive function has been shown to correlate highly with tests of crystallized function, and in the present study the genetic correlation with VNR was  $r_g = 0.783$  indicating that many of the same SNPs are involved in both facets of cognitive function. This overlap between these two phenotypes may therefore be driven by pleiotropic variants in introns and variants found in the H3K4Me1 histone mark, along with those that are evolutionarily conserved.

The strengths of this study include the use of the largest GWAS of general cognitive function that used established cognitive tests to measure cognitive function. We also use data from the UK Biobank study that includes over 30 000 participants genotyped and processed together using the same VNR test administered in an identical way to remove processing artefacts owing to heterogeneity in test used and their administration. In addition, the genetic data from UK Biobank were processed in a consistent manner.

The limitations of this study include the VNR test used in UK Biobank not being adequately compared to validated psychometric cognitive tests. Also the low response rate in UK Biobank of 5% (ref. 35) indicates that it may not be representative of the general population. A further limitation is the use of a general cognitive function phenotype derived from meta-analysis of many smaller studies. Heterogeneity in testing conditions and between different genotyping platforms could introduce a confound that could not have been controlled for here. However, it should be noted that these limitations would take away the power to find an effect should it be present and indicate the robust nature of the results. Future studies using more psychometrically rigorous tests and a sample drawn from across the population are therefore expected to provide additional insights into the genetic architecture of cognitive function. There may also have been a small number of individuals who took part in both CHARGE and UK Biobank. In addition, the stratified LD score regression method is based on an additive model and cannot detect epistatic effects or other sources of non-additive variance. Finally, as with other methods of GSA, these methods are limited to the availability and accuracy of the annotations used.

Following partitioned heritability analysis, we report that regions of the genome under negative selective pressure make a greater contribution to the heritability of cognitive functions than their size would suggest. This indicates that causal alleles are

not distributed evenly across the genome but, rather, cluster in regions that are conserved. Disease states and anthropometric traits that show genetic correlations with cognitive function tend to show enrichment in these conserved regions; on the other hand, the diseases and traits that do not show genetic correlations with cognitive function tend not to show this pattern of enrichment. Together, this suggests that conserved regions may have a central role in mammalian genetic architecture as they appear to harbour variants with pleiotropic effects between cognitive function, diseases and anthropometric traits. This suggests that understanding the function and role of variants that are under negative selection will provide a greater understanding of cognitive function as well as how cognitive function is related to other health-related traits.<sup>23</sup> That the variants here are known to be conserved within the mammalian line may also be involved in general cognitive function in other species.<sup>36</sup> In addition, this study aids the search for plausible sets of causal variants by showing that a reduced portion of the genome comprising, only ~2.5% of the total number of SNPs, can explain around ~40% of the SNP-based heritability (which is ~30%) of cognitive function.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGMENTS

The work was undertaken in The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross-council Lifelong Health and Wellbeing initiative (MR/K026992/1). This is supported by funding from the Biotechnology and Biological Sciences Research Council (BBSRC), the Medical Research Council (MRC) and the University of Edinburgh, and supports UD. This research has been conducted using the UK Biobank Resource under UK Biobank application 10279. The CHARGE Cognitive GWAS was supported by multiple grants including grants from the NIA (AG033193, AG049505). WDH is supported by a grant from Age UK (Disconnected Mind Project). UK Biobank received ethical approval from the Research Ethics Committee (REC reference 11/NW/0382).

## THE NEUROCHARGE COGNITIVE WORKING GROUP

Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium (CHARGE) cognitive working group banner includes: Gail Davies<sup>8,9</sup>, Ian J Deary<sup>10,11</sup>, Stephanie Debette<sup>12,13,14</sup>, Carla I Verbaas<sup>15,16</sup>, Jan Bressler<sup>17</sup>, Maaike Schuur<sup>18,19</sup>, Albert V Smith<sup>20</sup>, Joshua C Bis<sup>21</sup>, David A Bennett<sup>22</sup>, M Arfan Ikram<sup>23,24,25,26</sup>, Lenore J Launer<sup>27</sup>, Annette L Fitzpatrick<sup>28</sup>, Sudha Seshadri<sup>29,30</sup>, Cornelia M van Duijn<sup>31,32</sup>, Thomas H Mosley Jr<sup>33</sup>

<sup>8</sup>Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK, <sup>9</sup>Department of Psychology, University of Edinburgh, Edinburgh, UK, <sup>10</sup>Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK, <sup>11</sup>Department of Psychology, University of Edinburgh, Edinburgh, UK, <sup>12</sup>Department of Neurology, Boston University School of Medicine, Boston, MA, USA, <sup>13</sup>Institut National de la Santé et de la Recherche Médicale (INSERM), U897, Epidemiology and Biostatistics, University of Bordeaux, Bordeaux, France, <sup>14</sup>Department of Neurology, Bordeaux University Hospital, Bordeaux, France, <sup>15</sup>Genetic Epidemiology Unit, Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands, <sup>16</sup>Department of Neurology, Erasmus University Medical Center, Rotterdam, The Netherlands, <sup>17</sup>Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA, <sup>18</sup>Genetic Epidemiology Unit, Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands, <sup>19</sup>Department of Neurology, Erasmus University Medical Center, Rotterdam, The Netherlands, <sup>20</sup>Icelandic Heart Association, Kopavogur, Iceland. Faculty of Medicine, University of Iceland, Reykjavik, Iceland,

<sup>21</sup>Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA, <sup>22</sup>Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA, <sup>23</sup>Department of Neurology, Erasmus University Medical Center, Rotterdam, The Netherlands, <sup>24</sup>Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands, <sup>25</sup>Netherlands Consortium for Healthy Ageing, Leiden, The Netherlands, <sup>26</sup>Department of Radiology, Erasmus University Medical Center, Rotterdam, The Netherlands, <sup>27</sup>Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Bethesda, MD, USA, <sup>28</sup>Department of Epidemiology, University of Washington, Seattle, WA, USA, <sup>29</sup>Department of Neurology, Boston University School of Medicine, Boston, MA, USA, <sup>30</sup>The National Heart Lung and Blood Institute's Framingham Heart Study, Framingham, MA, USA, <sup>31</sup>Genetic Epidemiology Unit, Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands, <sup>32</sup>Netherlands Consortium for Healthy Ageing, Leiden, The Netherlands, <sup>33</sup>Department of Medicine and Neurology, University of Mississippi Medical Center, Jackson, MS, USA

## REFERENCES

- Carroll JB. *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. Cambridge University Press: New York, NY, USA, 1993.
- Davies G, Tenesa A, Payton A, Yang J, Harris SE, Liewald D et al. Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Mol Psychiatry* 2011; **16**: 996–1005.
- Marioni RE, Davies G, Hayward C, Liewald D, Kerr SM, Campbell A et al. Molecular genetic contributions to socioeconomic status and intelligence. *Intelligence* 2014; **44**: 26–32.
- Davies G, Armstrong N, Bis JC, Bressler J, Chouraki V, Giddaluru S et al. Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N = 53 949). *Mol Psychiatry* 2015; **20**: 183–192.
- Deary IJ. Intelligence. *Annu Rev Psychol* 2012; **63**: 453–482.
- Hill WD, Davies G, The CHARGE Cognitive Working Group, Liewald DC, McIntosh AM, Deary IJ. Age-dependent pleiotropy between general cognitive function and major psychiatric disorders. *Biol Psychiatry* 2015; **80**: 266–273.
- Hagenaars SP, Harris SE, Davies G, Hill WD, Liewald DC, Ritchie SJ et al. Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (N = 112 151) and 24 GWAS consortia. *Mol Psychiatry* 2016; **21**: 1624–1632.
- Davies G, Marioni RE, Liewald DC, Hill WD, Hagenaars SP, Harris SE et al. Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N = 112 151). *Mol Psychiatry* 2016; **21**: 758–767.
- Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet* 2012; **90**: 17–24.
- Wang K, Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. *Nat Rev Genet* 2010; **11**: 843–854.
- Hill WD, Davies G, van de Lagemaat LN, Christoforou A, Marioni RE, Fernandes CPD et al. Human cognitive ability is influenced by genetic variation in components of postsynaptic signalling complexes assembled by NMDA receptors and MAGUK proteins. *Transl Psychiatry* 2014; **4**: e341.
- Hill WD, de Leeuw C, Davies G, Liewald DCM, Payton A, Craig LC et al. Functional gene group analysis indicates no role for heterotrimeric G proteins in cognitive ability. *PLoS One* 2014; **9**: e91690.
- Holmans P, Green EK, Pahwa JS, Ferreira MA, Purcell SM, Sklar P et al. Gene ontology analysis of GWA study data sets provides insights into the biology of bipolar disorder. *Am J Hum Genet* 2009; **85**: 13–24.
- Mooney MA, Nigg JT, McWeeney SK, Wilmot B. Functional and genomic context in pathway analysis of GWAS data. *Trends Genet* 2014; **30**: 390–400.
- Segrè AV, Groop L, Mootha VK, Daly MJ, Altshuler D et al. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet* 2010; **6**: e1001058.
- de Leeuw CA, Neale BM, Heskes T, Posthuma D. The statistical properties of gene-set analysis. *Nat Rev Genet* 2016; **17**: 353–364.

- Lips E, Kooyman M, de Leeuw C, van Bochoven A, Posthuma D. JAG: a tool to evaluate the role of gene-sets in complex traits. *Genes* 2015; **6**: 238–251.
- Dixon L, Walter H, Schneider M, Erk S, Schäfer A, Haddad L et al. Identification of gene ontologies linked to prefrontal–hippocampal functional coupling in the human brain. *Proc Natl Acad Sci USA* 2014; **111**: 9657–9662.
- Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* 2015; **47**: 1228–1235.
- Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science* 2012; **337**: 1190–1195.
- Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA* 2009; **106**: 9362–9367.
- Schork AJ, Thompson WK, Pham P, Torkamani A, Roddey JC, Sullivan PF et al. All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. *PLoS Genet* 2013; **9**: e1003449.
- Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet* 2015; **47**: 12236–11241.
- Gusev A, Lee SH, Trynka G, Finucane H, Vilhjálmsson BJ, Xu H et al. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am J Hum Genet* 2014; **95**: 535–552.
- Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* 2016; **533**: 539–542.
- Penke L, Denissen JJ, Miller GF. The evolutionary genetics of personality. *Eur J Pers* 2007; **21**: 549–587.
- Deary IJ. Looking for 'System Integrity' in cognitive epidemiology. *Gerontology* 2012; **58**: 545–553.
- Bayés A, Collins MO, Croning MD, van de Lagemaat LN, Choudhary JS, Grant SG. Comparative study of human and mouse postsynaptic proteomes finds high compositional conservation and abundance differences for key synaptic proteins. *PLoS One* 2012; **7**: e46683.
- Bayés A, van de Lagemaat LN, Collins MO, Croning MD, Whittle IR, Choudhary JS et al. Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nat Neurosci* 2010; **14**: 19–21.
- Johnson MR, Shkura K, Langley SR, Delahaye-Duriez A, Srivastava P, Hill WD et al. Systems genetics identifies a convergent gene network for cognition and neurodevelopmental disease. *Nat Neurosci* 2015; **19**: 223–232.
- Jinks J, Eaves L. IQ and inequality. *Nature* 1974; **248**: 287–289.
- Jinks JL, Fulker DW. Comparison of the biometrical, MAVA, and classical approaches to the analysis of the human behavior. *Psychol Bull* 1970; **73**: 311.
- Joshi PK, Esko T, Mattsson H, Eklund N, Gandin I, Nutile T et al. Directional dominance on stature and cognition in diverse human populations. *Nature* 2015; **523**: 459–462.
- Marioni RE, Penke L, Davies G, Huffman JE, Hayward C, Deary IJ. The total burden of rare, non-synonymous exome genetic variants is not associated with childhood or late-life cognitive ability. *Proc R Soc B Biol Sci* 2014; **281**: 20140117.
- Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015; **12**: e1001779.
- Burkart JM, Schubiger MN, van Schaik CP. The evolution of general intelligence. *Behavioral and Brain Sciences* 2016; 1–65; doi.org/10.1017/S0140525X16000959.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016

Supplementary Information accompanies the paper on the Translational Psychiatry website (<http://www.nature.com/tp>)